DOCUMENT RESUME

ED 306 247 TM 013 020

AUTHOR Baker, Otis; And Others
TITLE The Missouri Comprehensive Statewide Project for Improving Student Achievement.
PUB DATE 89
NOTE 34p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; Achievement Gains; Achievement Tests; Criterion Referenced Tests; Elementary Secondary Education; Formative Evaluation; History; *Instructional Improvement; Learning Strategies; *Mastery Tests; *State Programs; *Test Construction; Testing Programs
IDENTIFIERS *Missouri Mastery and Achievement Tests

ABSTRACT

The history of a statewide effort to improve student learning in Missouri by adopting the concept of an Instructional Management System (IMS) with desired learned outcomes determining both instruction and testing was suggested to local school districts in 1980. In 1985, legislation provided funds and authority to engage large groups of teachers in creating criterion-referenced tests in reading and language arts, mathematics, science, and social studies and civics for various age groups. More than 800 educators were involved in developing Missouri's key skills. Over 200 teachers and administrators were involved in item writing, while another group of over 200 reviewed items for agreement with the key skills and for quality. With the help of consultants and further teacher review, field test forms were prepared and administered in April 1986. The original four forms were replaced by three revisions for an additional field trial conducted in 1987. The final version of the Missouri Mastery and Achievement Test (MMAT) batteries consists of 34 criterion-referenced tests assessing student performance in grades 2 through 10. Results of the 1988 administration of the MMAT document large achievement gains for Missouri students. Thirteen tables present data from test development. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED306247

The Missouri Comprehensive Statewide Project for Improving Student Achievement

ABSTRACT

This paper describes a statewide sustained effort to improve student learning by adopting and adapting promising research-based educational practices. The concept of an Instructional Management System (IMS) with desired learner outcomes determining both instruction and testing was suggested to local school districts beginning in 1980. In 1985, legislation provided both funds and authority to engage large groups of teachers to create criterion-referenced tests in reading/language arts, mathematics, science and social studies/civics for various age groups. The refinement process which resulted in the Missouri Mastery and Achievement Test batteries is described. The achievement gains of large populations of students as measured during the first year for which comparative data were available are also reported.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Richard L. King

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

We thank James H. Block, S. Alan Cohen, and Thomas R. Guskey for their comments on an earlier version of this paper. The opinions expressed and shortcomings that still exist, however, are attributable only to the authors.

TM013020

The Missouri Comprehensive Statewide Project for Improving Student Achievement

Otis Baker
Missouri Department of Elementary and Secondary Education

Richard King
Missouri Department of Elementary and Secondary Education

Kathleen M. Wulf
School of Education
University of Southern California-Los Angeles

Probably no other human enterprise, with the possible exception of medicine, has been researched more thoroughly than schooling. Ironically, no human enterprise has resisted change based upon research findings more than the schools. It has been speculated that if Rip Van Winkle had slept until the present day, the only place he would feel "at home" would be in the average U.S. public school classroom.

The Missouri Department of Elementary and Secondary Education (DESE) has decided over the last nine years to attempt to reverse this sad state of affairs. In an effort to focus a long term education improvement effort, the Department studied the available research to determine what interventions or practices contribute most to increasing the "effect size" of achievement gains. Drawing on the research data pertaining to curriculum alignment, instructional alignment (Cohen, 1986, 1987), mastery learning (Guskey, 1985; Block, 1988), and cooperative learning (Johnson and Johnson, 1984, 1987) we have formulated a research-based, outcome-based educational system which we call Instructional Management System (IMS).

The IMS concept is based on a large body of educational research, which can be summarized in the following two premises:

1. All students can learn what we teach to a satisfactory level--if teachers believe they can and if the school is organized to provide adequate time for each student to learn. (Carroll, 1963; Bloom, 1976)
2. For students to be self-motivated to continue learning throughout their school careers and through life, they must spend more than half their time working on tasks at which they experience a high degree of success. This implies that each student should encounter learning tasks at a rate no faster than he/she can master. Learning deficits should not be allowed to accumulate. (Ames & Ames, 1984; Guskey, 1985)

In June, 1981 the Missouri State Board of Education adopted as a top priority the implementation of an IMS in every district of the state. The remarkable feature of this action was that it made IMS a primary Board priority for the "decade of the 80's." Seldom do policymaking boards understand the time required to bring about substantial change--especially when the constituency is large and the change is not mandatory. (Mo. DESE, 1986)

During the summer of 1981 the Department staff conducted six day-long conferences on IMS in Jefferson City, the state capitol. A total of about 800 school personnel, mostly administrators, attended these meetings. They were introduced to the IMS concept and informed of the renewed state commitment to outcome-based schooling. Their response ranged from enthusiasm to skepticism, both toward the idea of IMS and toward the long-term state commitment to significant change. One superintendent, who was asked by a colleague how he planned to respond to the call for an IMS, said, "I'm just going to hunker down until it goes away."

During this same time, two or three of the Department staff members conducted meetings at all the state colleges and universities. The purpose of these meetings was to explain to teacher educators the IMS concepts and to elicit their support by incorporating these concepts into teacher and/or administrator training programs. We also held a day-long meeting in Jefferson City for teacher educators to inform them of our proposed emphasis

on IMS. This meeting was attended by 87 college and university professors.

These efforts have had mixed results. Relatively few teacher educators have provided thorough training in outcome-based or mastery learning practice; still fewer have incorporated these concepts into their own classes. Teachers, for the most part, find the development of an IMS or organizing for mastery learning quite difficult. The demand for technical assistance is high. Even teachers who have recently completed their training (whether in Missouri institutions or in other states) feel inadequate when it comes to instructional alignment. They are unaware of the subtleties of directing instruction toward the outcome assessment. This has not been demanded of them nor taught in teacher training. Teachers' experience and training has often resulted in expectations that instruction causes a normal distribution in student achievement when the exact opposite of this is a major premise of an IMS.

To address this need for inservice training, Department staff have conducted three or four week-long IMS workshops each summer at state universities for graduate credit. Enrollment at each workshop has been limited to 100. Applicants exceeding this number have been turned away at nearly every scheduled workshop. Most of the workshop participants have come as local school district teams. Enrollment fees, paid by local districts or by individuals, have paid the costs of these workshops. In addition, during the school years since 1981, many local district staff development workshops have been conducted by the Department staff.

The Board and Department staff have also tried to build public awareness and understanding of the IMS concept. In recent years, for example, the Department has conducted an annual series of 10 "Regional Educational Conferences" in order to provide a forum for public discussion of a current educational issue. Each fall these meetings have attracted about 5,000 Missourians, including citizens, school board members, community leaders, students, legislators and school district personnel.

The meetings include large-group presentations by state education officials, small-group discussion sessions for

participants, and an opportunity for participants to express their views on current topics. During 1981 and 1982, the Department's Regional Conferences focused on instructional management and outcome-based education. The same topic was featured two consecutive years to emphasize the importance of these issues and to demonstrate a state-level commitment to the goal of promoting outcome-based education in all public schools.

Further impetus toward an outcome-based approach to learning was provided by the 83rd Missouri General Assembly. In 1985 the legislature passed House Bill 463, the "Excellence in Education Act of 1985." This Act dealt with all aspects of education simultaneously from teacher training and certification to curriculum standards and testing for students. Following is the complete text of Section 4 of the Act, which has provided a framework whereby efforts to implement an IMS could be accelerated across the state:

H.B. 463

SECTION 4

(Emphases added.)

1. The commissioner of education shall direct the department of elementary and secondary education to insure that all school districts have a program of pupil testing which shall test competency in the subject areas of English, reading, language arts, science, mathematics, social studies and civics.
2. The department of elementary and secondary education <u>shall identify key skills</u> within the subject areas contained in subsection 1 of this section which shall provide the foundation for the local school district's testing program. The department of elementary and secondary education may not set maximal testing standards.
3. Each local school district shall have a testing program. District testing programs may include minimum promotion standards and shall give due consideration to the research on the influence of cultural diversity on testing performance.

4. The testing program of each local school district shall include, but shall not be limited to, criterion-referenced tests approved by the department of elementary and secondary education. This testing program shall test all students at periodic grade levels. The testing program may test students and shall identify areas for instructional improvement. The department of elementary and secondary education may develop criterion-referenced tests and assist districts with their testing programs upon the district's request.
5. The department of elementary and secondary education shall develop or select tests which measure student performance on minimum key skills, and shall annually administer such tests to a randomly selected, statewide sample of public school students.
6. Each local school district shall provide testing information upon request to the department of elementary and secondary education.
7. The department of elementary and secondary education shall annually report to the general assembly composite pupil testing information.

Developing Core Competencies/Key Skills

As soon as it became apparent that the Excellence in Education Act would pass and be signed into law, the Commissioner of Education appointed an Oversight Committee. This committee consisted of more than 100 persons representing all professional education organizations, labor leaders, agriculture and business leaders and legislators. After sharing DESE proposed plans for implementing Section 4 of H.B. 463 and receiving several suggestions from members of the Oversight Committee, statewide committees of teachers, subject area specialists and school administrators were appointed to recommend the "core competencies and key skills" to meet the mandate of the law.

Several critical decisions were made before the key skills were selected:

1. The decision was made to start at grade 10. This is generally the end of the "common" core curriculum experienced by all students. In addition, we decided to list as "key" those skills in each subject area which would undergird any elective courses a student might choose at grades 11 and 12. Finally, the tenth grade was selected because most of the high school dropouts in Missouri occur during grades 11 and 12. Thus, we could test a greater percentage of students than if we waited until grade 12 to test.
2. We decided to "work down" from grade 10. What does a student need in grade 9 as prerequisite to doing well in grade 10? What in grade 8?, etc.
3. Another decision was made to go only as low as grade 2 in the list of core competencies and key skills. Grade 1 skills were to be listed along with those for children ages four through six. These are now being developed.
4. It was also decided that we should include important outcomes that cannot be measured by a paper/pencil test in the key skills list. These would, of necessity, be assessed locally.
5. Because the legislation calls for testing of a statewide sample of students at "periodic" grade levels, we chose to prepare tests for grades 3, 6, 8 and 10 first.
6. Student writing would be assessed each year on a sample of students at grades 6, 8 and 10. Holistic scoring would be done, for the state sample, by teachers who have been trained in this approach of assessing writing.
7. These tests would not be "grade level" tests in the usual sense of the word. They would not only measure what had been taught the year the test was taken, but also they would sample those skills, competencies and knowledge that a student had <u>accumulated</u> as of early April each year. This was done to spread accountability for student learning beyond the individual teacher; to include the building administrator/s. How well is the principal managing the resources in his/her building to assure

every student a quality school experience?

More than 300 educators, from primary grade teachers through college professors, were involved in developing Missouri's key skills. Our "Core Competencies and Key Skills" (formally published in the fall of 1986) do not include all that should be learned through grade 10. Instead, the state list is a skeleton for local districts to "flesh out" as they see fit. The skeleton analogy is rather appropriate since everyone's skeleton is very similar when viewed in an x-ray. It is only when the necessary flesh and blood are added that the individual personal characteristics make us individuals; this allows for much local curriculum autonomy. Also, a skeleton is connected from the "toe bone" to the "head bone;" similarly, our key skills are also vertically articulated.

Several other concepts that research has demonstrated add to the "effect size" of achievement gains have been incorporated into Missouri's plan for achievement gains:

1. The key skills were developed to build sequentially the concept of "enhanced prerequisites." (Leyton, 1983) If the teachers of sequential courses agree on the prerequisites needed for the subsequent course in a specific school subject, then each teacher of the first course can insure that by the end of the term most of the students have mastered these prerequisites before going on to the subsequent course in the subject.
2. The importance of home participation and support and that of enhanced cues and reinforcement through frequent feedback are stressed. Learner outcomes and mastery criteria are clearly spelled out before instruction begins. Both the student and his/her parents are provided this information. After initial instruction and the first formative test the student is provided feedback as to what he has learned well and what he still needs to work on. This coupled with the mastery learning process has been shown to have a possible effect size of two sigma. (Tenenbaum, 1982)
3. Teaching for higher order thinking has been shown to add to the total effect size, also. (Mevarich, 1980)

> Cognitive psychologists have classified more than half of Missouri's core competencies/key skills as being above the level of "comprehension" on Bloom's Taxonomy of Educational Objectives. (Bloom, 1956)

The Missouri Department of Elementary and Secondary Education sponsored two major statewide conferences: "Beyond Effective Schools" in 1986 and "Beyond Effective Schools II" in 1987. These were planned to let school district teams learn more about mastery learning, cooperative learning and OBE. Major presenters at these conferences have been Carol Barber, James Block, John Champlin, Alan Cohen, Thomas Guskey, David Johnson and Susan Loucks-Horsley. Care was taken to have each of these presenters come to the Department for pre-conference consultation to insure that the focus of the conference was sharp and clear. Several of these speakers have returned to the State after the meetings to provide workshops in their area of expertise.

Draft copies of the proposed Core Competencies and Key Skills were distributed and discussed at fourteen regional curriculum conferences across the state which were attended by more than 4000 teachers and administrators during the Fall, 1985. Suggestions obtained at these meetings were incorporated in the final draft of the skills.

As the Core Competencies/Key Skills were being refined the DESE met with representatives from every major test publisher in the United States to see if they could get tests developed which would meet Missouri's needs. There were two important reasons for DESE to reject all proposals by commercial testing companies to produce the tests: first, the items in their item banks which they showed DESE were almost all at the level of simple recall or memory and, second, the testing company representatives were so indoctrinated with the psychometric model of testing (normal curve as a design target) that they all proposed embedding a norm referenced set of items in the tests. The first design shortcoming violated the concept of curriculum and instructional alignment and the second was not consistent with the goal of mastery for all students. The DESE was committed to <u>developing</u> talent as opposed to selecting talent.

Developing the Tests

When the Department decided to develop its own tests, which were to be called the Missouri Mastery and Achievement Tests (MMAT), a contract was negotiated with the Center for Educational Assessment (CEA) at the University of Missouri, Columbia to assist in organizing, developing, refining and finally, providing test scoring and reporting strategies.

The MMAT was developed in two phases. Phase I, grades 3, 6, 8, and 10, was developed from 1985 through 1987. Phase II, grades 2, 4, 5, 7, and 9, was developed from 1986 through 1988. Virtually identical development processes were used in the two phases. The tests for grades 3, 6, 8, and 10 were first administered in 1987, while the tests for grades 2, 4, 5, 7, and 9 were first administered in 1988.

Insert Figure 1 about here

A test development effort such as this one requires large numbers of individuals to complete specified tasks at each stage of test development. While many individuals were involved in developing the MMAT, the groups described below were formally identified to perform the functions indicated.

Widespread involvement of Missouri educators ensured that the tests were crafted with a firsthand knowledge of the state's school children and a special sensitivity to the educational growth of its young people. After the test objectives were determined, creation of the test battery itself required contributions by hundreds of Missouri educators at every level and scrutiny and revision of numerous drafts by CEA and the Missouri Department of Elementary and Secondary Education.

More than 200 teachers and administrators from schools throughout Missouri were involved in item writing. Another group of more than 200 teachers reviewed the items written by their colleagues for agreement with the Key Skills and for quality of the item stem and its choices. Teachers and administrators were involved in every phase of test development. In addition, a

sample of students who had participated in the field trials was interviewed with structured-interview protocols to assess which features of the items had prompted their responses.

Consultants from the Department of Elementary and Secondary Education were involved with test development at all stages. Department consultants worked with CEA staff, teachers, and administrators in defining test content and in item writing.

The CEA staff supplemented its expertise with numerous consultants who worked on specific aspects of test development. A group of 12 nationally noted college professors provided special expertise in curriculum, content, and learning theory. This group worked initially to define the test content specifications and then reviewed the item development as the process progressed. A panel of five testing experts from across the country served as the Technical Advisory Committee and provided input on issues concerning test construction, scaling, equating, norming, item analysis, reliability, and validity. Members of that panel have participated individually in completing specialized tasks and studies during test development. A nationally recognized specialist in sampling design devised the sampling plans employed in this effort. A number of individuals examined items for bias. In short, CEA utilized the skills of no fewer than 20 nationally recognized individuals to amplify its professional efforts. Finally, since state officials and leaders in business, industry, and education were involved as an Oversight Committee during the implementation of the Excellence in Education Act of 1985, the implementation of the Act was addressed by utilizing the services of the broadest sampling of the community involved in education.

A committee composed of 12 professors, three each from reading, social studies, mathematics, and science, wrote content specifications for test items for each of the 688 skills in both phases of development (Phase I, 353; Phase II, 335). Particular skills were assigned according to an individual's content or curricular expertise. The professors were instructed to elaborate, interpret, and clarify each key skill so that its particular characteristics might be clearly and uniformly understood by everyone involved. These elaborations formed the initial draft of the test content specifications.

Practitioners Develop the Test Items

Next, approximately 120 teachers from all grades throughout Missouri were assembled to review and further refine the specifications drafted by the college content area specialists. Test content specifications writing workshops were convened for both phases: December 4-5, 1985, for Phase I; and September 7-9, 1986, for Phase II. Workshop participants were charged with reviewing specifications in terms of their applicability to classroom instruction. For each phase, teachers met in 16 groups; a group of five to seven teachers met for each grade level and for each content area. Phase I groups reviewed reading and English/language arts, mathematics, science, and social studies specifications for grades 3, 6, 8, and 10. Phase II groups reviewed specifications for grades 2, 4, 5, 7, and 9. Curriculum consultants from the Department of Elementary and Secondary Education assisted the teachers in reviewing specifications. The reviewers' comments and the exemplary items they produced were compiled to describe each skill according to content domain and grade level.

This version of the test content specifications was then reviewed by the college content area specialists. CEA editorial staff and Department consultants additionally reviewed and revised the specifications. Finally, each test content specification was reviewed by Department consultants and two to five teachers for specific content at a particular grade level. In total, more than 1,000 reviews of test content specifications had occurred by the time they reached their final form. The final test content specifications were then printed and distributed to all school districts in Missouri. The complete document, Test Content Specifications, Grade 2-10, is available from CEA. Many teachers use the specifications in preparing lesson plans.

A test item writing workshop was conducted from January 6 through 10, 1986, for Phase I. For Phase II separate grade level workshops were held from September through November 1986. As with the Test Specifications Conference, teachers were grouped according to their subject areas and grade levels. Test specifications

for the grade level and the content area were available to each group.

Two test item construction experts taught participants techniques of item writing during the workshops. In addition, consultants for the Department and from CEA worked closely with the teachers.

Hundreds of test items, obtained from public domain item banks, were available to item writers during the workshop. The teacher/authors decided that these items were not appropriate for assessing the Key Skills, and that the content of the items themselves was not suitable. The items were abandoned in favor of developing original items referenced directly to the test content specifications and Key Skills.

By the close of the workshops, 12 to 16 items for each of the 688 test content specifications had been developed. This translated to approximately 4,800 items for each phase, just under 10,000 items during the two phases.

In each of the two phases of test development, another group of approximately 120 teachers reviewed the items. Content specialists, including college professors, teachers, private consultants, curriculum consultants from the Department, and the CEA staff, reviewed the items. This process provided at least five reviews per item. Each systematic review required the use of a scorecard which included specific criteria relating to congruence between item and the key skill, appropriateness of the reading level for the grade level, wording, format, and appropriateness of distractors. These reviews were analyzed and the results were synthesized by CEA editorial staff and curriculum consultants from the Department of Elementary and Secondary Education. Modifications of particular items were made when appropriate. In some instances, these exhaustive reviews resulted in the removal of test items from further consideration.

Studies of the readability of items were completed by a reading specialist using standard formulas such as Flesch, Dale-Chall, SMOG, and FOG. For grades 2 and 3, a panel of 15 classroom teachers reviewed the test items and the directions to ensure that the syntax and vocabulary would be understood by children of this age.

For both phases, item bias reviews, both judgmental and empirical, were undertaken. A panel with nationally accepted expertise in test item bias judged items for bias. Approximately 150 classroom teachers were asked to rate items systematically with specially constructed rating sheets to point out stereotyping or other objectionable wording which might bias items. Each item was rated approximately five times by classroom teachers who examined how each item might affect the way a child with a culturally or linguistically different background might respond. For each bias review, clearly delineated criteria were used. Items identified as biased were either discarded or flagged as needing editorial repair. Additionally, CEA editorial staff adjusted items to ensure a mix of genders in the name references in items, a mix of traditional and non-traditional gender roles, and a variety of ethnic backgrounds. Similar, though not identical, processes were used in both phases of test construction.

The initial test item pool contained approximately 9,600 items. The text alone for each phase required some 70,000 lines. More than 800 galleys were set. Five graphic artists completed approximately 1,500 graphics for the items for each phase. Just under one-third of the items required graphics, and a large number required text to accompany items.

Establishing Item and Alternative Form Reliability

An unedited version of the test was produced in March 1986 during Phase I. Four forms were produced so that the staff of CEA could evaluate how the items looked in whole test format. These four forms were only for a preliminary review of items and were not used for testing with students. However, they served as the starting point for the finished product.

During Phase I four field test forms were created: Forms W, X, Y, and Z. Items were assigned to forms by these rules:

1. Form W would always have four items per key skill. The remaining items for Key Skills would be distributed across Forms X, Y, and Z so that each had equivalent numbers of items, content, and complexity of items.

2. Items which might give clues to correct answers to other items in the pool would not be placed in the same form.
3. The number of items linked to a passage or a graphic would increase as the grade level increased.

A total of 64 field test booklets was constructed: four forms for four content areas for four grade levels. Each booklet went through 26 steps of editing, revision, and production to reach final distribution. Each question went through two rounds of editing with at least four persons reading each item. A data base was established for every item, referencing response options, graphics, and text to the item stem. All Key Skills were contained in each of the four forms. Four or more items were assigned to each Key Skill in every form. Graphics and art work were further edited to satisfy the demands of page design.

Field tests were administered during the week of April 28 to May 2, 1986, to a representative sample of school districts throughout the state. Geographic location, school size, and socio-economic status comprised criteria for selection within a sampling design. The field trial included approximately 3,500 students from about 350 schools in 250 school districts. Each form was administered at its designated grade level and at the grade level below and above to assure measurement of a broad range of abilities. In addition, the Iowa Tests of Basic Skills (ITBS) for grades 3, 6, and 8 and the Tests of Achievement and Proficiency (TAP) for grade 10 were administered to provide pilot-study data for obtaining Chapter 1 usable norms. Each student in the field trial sample was administered only one subject area of the MMAT and the corresponding area of the ITBS or TAP.

For the Phase II field trial, Forms M, N, and O were created using the rules and procedures previously described. These forms were administered from April 20 to 24, 1987. A sampling design similar to that of Phase I was employed, and numbers of students taking the test were virtually identical. Initial item response data were utilized to evaluate items in the field testing forms.

During Phase I about 400 items were chosen for review with students in the classroom during May 1986. These items were neither excellent nor poor; rather, they were in the middle group, an unknown quantity. For three weeks, members of the CEA staff

and Department consultants visited three to four classrooms a day throughout the state to discuss with students their opinions and reactions to particular test items. These in-class discussions were conducted at grades 3, 6, 8, and 10, and from them reviewers tried to determine the quality and validity of particular test items. Student responses to items were discussed with the students using a structured interview form. Each interviewer completed a protocol for every item discussed. Information gathered in this way was used to conduct a qualitative analysis of some of the items. Teams of two teachers and one Department consultant reviewed items during June 1986 for Phase I and during June 1987 for Phase II, using Key Skills, test content specifications, teacher review comments, item statistics from the field testing, and classroom review data.

This exhaustive process of item development, review, and revision has resulted in a pool of items of exceptional quality. The language of each item had been clarified; its congruence to a specific skill had been verified, and its ability to be understood by students had been examined. Several of the guiding principles which insured the quality of the resulting test include:

- Field participation. Teachers and subject area consultants established the content validity of the test.
- Expert judges. Curriculum and measurement authorities reacted to each stage of the test development and field trial techniques.
- A common-items equating design was used to ensure that the final forms for each level were indeed parallel. Forms A, B and C for grades 3, 6, 8 and 10 underwent a field trial during the fall of 1986 in order to obtain information about testing time, administration procedures, and equivalency of forms within a subject and level.

The final version of the MMAT consists of thirty-four criterion-referenced tests which assess student performance on Key Skills in grades 2 through 10. The grade 2 tests cover Language Arts/Reading and Mathematics. Subject tests for grades 3 through 10 include Reading/English/Language Arts, Mathematics, Science, and Social Studies/Civics. The battery consists of multiple-choice items with four options each. There are three equivalent test

forms (A, B, and C) for grades 3, 6, 8, and 10 and two equivalent test forms (D and E) for grades 2, 4, 5, 7, and 9. There are two parts to each of the subject tests; this allows students to be given a rest break during each subject test.

Insert Table 1 about here

Using the MMAT to Test School Improvement

In the Spring of 1987, approximately 240,000 students in grades 3, 6, 8, and 10 were administered Form A. In the Spring of 1988, approximately 240,000 students in grades 3, 6, 8, and 10 were administered Form B and approximately 200,000 students in grades 2, 4, 5, 7, and 9 were administered Form D. During the current school year more than 83 percent of Missouri's students are scheduled to take the MMAT in grades 2 through 10; the remainder are giving the test to the required minimum of four levels.

Each MMAT form consists of 34 district subject tests assessing student performance in grades 2 through 10. Tests for grade 2 cover two subjects:

1. Reading/Language Arts
2. Mathematics

Tests for grades 3 through 10 cover four subjects:

1. Reading (grades 3-6) or English (7-10)/Language Arts
2. Mathematics
3. Science
4. Social studies

The subject tests for each grade are bound together in a booklet. Grade 2 tests are presented in machine-scored booklets that have been especially designed to accommodate the requirements of students at that level and that eliminate the need for separate answer sheets. The booklets for grades 3 through 10 are color coded and have matching machine-scorable answer sheets.

The tests are administered in early April, scored, and the reports are returned to schools during the last two weeks in May; before school is out for the year.

Local school district reports include the following:

1. Individual student report.

   This report lists, by subject, the Key Skills mastered and those not mastered, the student's scale score (on a scale of 100-500) and the district average scale score for that grade and subject.

   The scale score of 300 was arbitrarily set to reflect the state average performance for 1987. A sample of schools representing rural/urban, large/small, high SES/low SES was randomly chosen (more than 5,000 students per grade level) and the same sampling process is used each year to obtain the State sample.

2. Individual Student Score Labels are supplied to be affixed to the student's permanent record.

   The ITBS and the TAP are administered each year to the State's Sample to provide comparable national percentile ranks. This also allows the State to provide Chapter I eligibility lists to each district. This label lists the student NCE.

3. Pupil List Report.

   This report lists all students, by grade and subject, in a school building with mastery/non-mastery of each Key Skill indicated.

4. Grade Level Key Skill Report.

   This report presents the percentage of students mastering each Key Skill. There is a separate Grade Level Key Skill Report for each subject for a grade in a building and in a district.

5. Grade Level Cluster Report.

   This report depicts in graphic form the performance of students in a specific building or a district. Each "x" corresponds to the average scaled score of the students in a building or the district at that grade on a subject or cluster. The bar extending through the "x" represents the standard deviation and shows the degree of

variation among the scores.

6. Summary Report.

   This report provides information about an entire building's or district's performance at grades 2 through 10 on each subject tested on the MMAT. The Summary Report, when compared to previous years' reports, can be a useful tool for monitoring changes over time in MMAT performances.

7. District Summary Report by Quintiles.

   Quintiles for the State sample scores were established in 1987. Since then school district quintile scores are reported upon request. This report displays what percent of students in a local district fall in each fifth of the state rankings established in 1987. This form is especially valuable to monitor progress toward equity in learner outcomes--which is a goal of OBE.

8. A school district may elect to receive other reports including summary data disaggregated by socio-economic status, race, or sex, as well as output on computer diskette or tape.

The results of the 1988 administration of the MMAT demonstrated that the Department's optimistic expectations were not misplaced. All of the state's sample showed gains. Obviously all of the more than 500 school districts in Missouri had not implemented all aspects of an IMS in all subjects and levels.

Insert Table 2 about here

The gains will more nearly approach what research has shown is possible as more districts adopt OBE practices.

Some indication of this can be seen in the performance of 185,000 students enrolled in 109 school districts which are implementing more of the OBE strategies and which expressed an interest in including learning output data in their annual school classification report. These districts, which had a total of 185,000 students showed gains ranging up to more than one sigma in the one year period.

Insert Table 3 about here

The ITBS and the TAP were also administered to the State sample students at grades 3, 6, 8 and 10. This was done, not only for obtaining normal curve equivalent data, but also to demonstrate that this statewide focus on a set of learner outcomes which were not constructed to match a norm referenced test did not cause a decline in NRT scores. It is interesting to note that the gain on the TAP in tenth grade science was from the 48th to the 64th percentile. This may lend support to the proposal of mastery learning proponents "that the real power of mastery learning should lie in its capacity to help teachers of open and unstructured subjects such as science to become clearer about their instructional objectives and more systematic in attaining them." (Block, 1988, p. 30)

Insert Table 4 about here

An even more crucial measure of how this entire process is bringing about improvement in the equity of student progress can be seen in the quintile reports. The State sample had exactly 20 percent of the students in each quintile in 1987 because that is how the quintiles were established. One year later we can see that a higher percent of students are in the higher one-fifth and fewer in the lower.

Insert Table 5 about here

The 109 "Academic Achievement Districts" made even more difference.

Insert Table 6 about here

There are two interesting observations about this table: These districts had nearly state average performance in 1987 and, second, they have quite high percents of students in the top two

fifths and remarkably lowered percents in the bottom fifth. This is after only one year!

When one looks at the state sample (more than 20,000 students each year) and see statistically significant gains at all four levels and in each four subjects, the question arises: What caused these gains? Mastery learning and cooperative learning were only common practice in fewer than half the districts of the state. The only changed practice which was common to all schools was curriculum and instructional alignment. The mean scores in 1988 had moved far enough from the mean established in 1987 to cause a notable increase in the standard deviation from one year to the next in all cases. (Table 1)

The insensitivity to instruction of the NRT (in this case the ITBS and TAP) is apparent in Table 4. If we want school improvement, we must monitor it; but we must use tests designed to detect improvement.

Using the MMAT to Cause School Improvement

As students and teachers become more adept in outcome based instructional practice and as students experience more mastery learning and cooperative learning we should expect to see continued substantial gains in school achievement. At present a revision of the Core Competencies and Key Skills is under consideration by four large committees of teachers and subject area supervisors. This will be followed by the development of additional forms of the MMAT. This will be an ongoing process.

Several lessons have been learned by Missouri's experiences in the quest for improved student learning:

1. Research-based practices which show <u>educationally significant</u> promise, such as outcome based educational decision making, mastery learning and instructional alignment, are the only practices that justify a major state effort.
2. The role of the State Board of Education and the Commissioner of Education is extremely important to successful statewide changes. Their long term public

commitment to instructional management has been crucial to the gains made in Missouri schools.

3. The credibility and practitioner ownership of a process is positively related to the breadth of involvement of those most affected by the process.
4. Many change efforts fail because the newness of the change wears off and because new ideas intrude before sufficient time is devoted to the original practice to see it successful. Focus over time is needed in staff development.

BIBLIOGRAPHY

Ames, R. E. & Ames, C. (Eds.) (1984). Research on motivation in education (Vol. 1): Student motivation. Orlando, FL: Academic Press.

Block, J. H. & Burns, R. B. (1976) Mastery learning. In L. Shulman (Ed.) Review of research in education (Vol. 4). Itaska, IL: Peacock.

Bloom, B. S., et al. (1956). Taxonomy of educational objectives: handbook 1, cognitive domain. New York: McGraw-Hill.

Bloom, B. S. (1968). Learning for mastery. (UCLA-CSEIP) Evaluation comment. 1-12.

Carroll, J. B. (1963). A model of school learning. Teachers College Record, 64, 723-733.

Cohen, S. A. (1987). Tests: marked for life? Ontario, Canada: Scholastic.

Cohen, S. A. (1988). Instructional alignment: searching for a silver bullet. Educational Researcher AERA.

Guskey, T. R. (1985). Implementing mastery learning. Belmont, CA: Wadsworth.

Johnson, D. W., & Johnson, R. T. (1985) Cooperation in the Classroom. Edina, MN

Johnson, D. W., & Johnson, R. T. (1987) Learning together and alone: competition, and Individualization. (2nd edition)

Lysakowski, R. S., & Walberg, H. J. (1982). Instructional effects of cues, participation, and corrective feedback: A quantitative synthesis. American Educational Research Journal, 19, 559-578.

Leyton, F. S. (1983). The extent to which group instruction supplemented by mastery of the initial cognitive prerequisites approximates the learning effectiveness of one-to-one tutorial methods. Ph.D. dissertation, University of Chicago.

Mevarech, Z. R. (1980). The role of teaching-learning strategies and feedback and corrective procedures in developing higher cognitive achievement. Unpublished doctoral dissertation, University of Chicago.

Missouri Department of Elementary and Secondary Education (1989). Missouri mastery and achievement tests technical manual. Columbia, MO: Center for Educational Assessment.

Missouri Department of Elementary and Secondary Education (rev. 1986). Introduction to instructional management (IMS), Jefferson City, MO: Division of Instruction.

Nordin, A. B. (1979). The effects of different qualities on selected cognitive, affective, and time variables. Ph.D. dissertation, University of Chicago.

Soled, S. W. (1986). Group instructional methods which are a effective as one-to-one tutoring in the improvement of both higher as well as lower mental process achievement. Ph.D. dissertation, University of Chicago.

SUBJECTS AND CLUSTERS

Reading or English/ Language Arts	Mathematics	Science	Social Studies/ Civics
Language Arts	Understanding Numbers	Life Science	Geography
Reading	Computation	Earth Science	History
Writing	Measurement and Geometry	Physical Science	Government
	Interpretation and Application	Cross-Disciplinary	Economics
			Civics

Figure 1 -- Clusters of key skills included in the Missouri Core Competencies/Key Skills by subject area

Grade 3, Form A and Form B Statistics
Spring 1987/1988 Administrations

Subject	Number of Items	Scaled Score: Means Forms: A	B	Standard Deviation Forms: A	B	IRT Reliability Forms: A	B
Reading	84	300	315*	65	68	.948	.941
Mathematics	68	300	318*	65	77	.929	.916
Science	64	300	311*	65	72	.909	.895
Social Studies	56	300	325*	65	73	.924	.905

n = 4789 5706

Grade 6, Form A and Form B Statistics
Spring 1987/1988 Administrations

Subject	Number of Items	Scaled Score: Means Forms: A	B	Standard Deviation Forms: A	B	IRT Reliability Forms: A	B
Reading	92	300	308*	65	72	.937	.937
Mathematics	104	300	332*	65	77	.955	.956
Science	92	300	332*	65	84	.914	.926
Social Studies	84	300	321*	65	77	.943	.942

n = 5138 5152

Grade 8, Form A and Form B Statistics
Spring 1987/1988 Administrations

Subject	Number of Items	Scaled Score: Means Forms: A	B	Standard Deviation Forms: A	B	IRT Reliability Forms: A	B
Reading	108	300	322*	65	70	.951	.950
Mathematics	100	300	336*	65	84	.949	.953
Science	72	300	337*	65	88	.901	.906
Social Studies	72	300	325*	65	77	.941	.937

n = 5346 4733

Grade 10, Form A and Form B Statistics
Spring 1987/1988 Administrations

Subject	Number of Items	Scaled Score: Means Forms: A	B	Standard Deviation Forms: A	B	IRT Reliability Forms: A	B
Reading	112	300	315*	65	72	.948	.940
Mathematics	92	300	313*	65	72	.951	.952
Science	80	300	326*	65	82	.891	.892
Social Studies	100	300	314*	65	70	.957	.952

n = 5645 5023

*All gains are significant at the .0001 level.

Table 1 -- The standard error range is from .9 to 1.2

MISSOURI MASTERY AND ACHIEVEMENT TESTS

ACADEMIC ACHIEVEMENT REPORT

State Sample: 1987 and 1988

Subject / Grade	READING		MATH		SCIENCE		SOCIAL ST.	
	87	88	87	88	87	88	87	88
3	300*	315	300	318	300	311	300	325
6	300	308	300	332	300	332	300	321
8	300	322	300	336	300	337	300	325
10	300	315	300	313	300	326	300	314

* -- Sigma for the baseline (300) is approximately 65.

Table 2 -- State sample scale scores for 1987 and 1988. (n = 5000 students at each level)

Academic Achievement Districts: 1987 and 1988

Subject / Grade	READING		MATH		SCIENCE		SOCIAL ST.	
	87	88	87	88	87	88	87	88
3	306 n = 14,093	334 n = 14,793	304 n = 14,236	349** n = 14,818	307 n = 14,228	335 n = 14,806	314 n = 14,178	340 n = 14,746
6	298 n = 13,409	323 n = 13,188	303 n = 13,433	379* n = 13,178	305 n = 13,462	393* n =13,165	304 n = 13,474	358** n = 13,147
8	310 n = 13,568	349** n = 13,198	317 n = 13,647	387** n = 13,225	314 n = 13,621	385** n = 13,211	314 n = 13,629	362** n = 13,179
10	304 n = 14,890	327 n = 13,501	309 n = 14,957	328 n = 13,525	308 n = 14,958	356** n = 13,561	305 n = 14,900	329 n = 13,467

* one sigma gain, sigma for the baseline (300) is approximately 65.

** gains from .56 to .83 sigma

Table 3 -- Scale score gains recorded for 109 Missouri school districts which volunteered to use achievement data as a school classification standard. These districts have total enrollments of 185,000 students.

Median Percentile Scores of Students in State Sample on ITBS (grades 3, 6, 8) & TAP (grade 10)

Grade / Subject	Grade 3		Grade 6		Grade 8		Grade 10	
	87	88	87	88	87	88	87	88
Reading	56	56	50	60	50	52	53	51
Language Arts	64	64	49	53	45	50	--	--
Math	55	59	55	67	59	64	54	59
Science	65	68	65	65	63	70	48	64
Social Studies	61	70	58	60	52	64	65	65

Table 4 -- Median percentile scores of students in the state sample on ITBS, grades 3, 6, 8, and TAP grade 10. (n = about 5000 students at each level)

MISSOURI MASTERY AND ACHIEVEMENT TESTS

ACADEMIC ACHIEVEMENT REPORT

State Sample 1988

GRADE 3

	READ	MATH	SCI	SOC. ST
HIGH	28	30	26	32
	21	20	23	24
	18	18	16	18
	17	15	17	13
LOW	15	17	18	13
Number Tested	5706	5706	5706	5706

GRADE 6

	READ	MATH	SCI	SOC. ST
HIGH	24	36	37	32
	19	21	19	20
	19	17	15	16
	19	14	14	16
LOW	18	12	15	15
Number Tested	5152	5152	5152	5152

GRADE 8

	READ	MATH	SCI	SOC. ST
HIGH	31	40	38	34
	22	18	20	20
	18	16	15	17
	16	14	13	14
LOW	13	13	14	15
Number Tested	4733	4733	4733	4733

GRADE10

	READ	MATH	SCI	SOC. ST
HIGH	27	27	34	26
	21	21	20	22
	18	18	16	20
	17	17	13	17
LOW	16	17	16	16
Number Tested	5023	5023	5023	5023

Table 5 -- 1988 percent of students scoring in each fifth established using 1987 data.

Academic Achievement Districts: 1987 and 1988

GRADE 3

	READ		MATH		SCI		SOC. ST	
	87	88	87	88	87	88	87	88
HIGH	23	40	25	45	24	38	27	38
	21	22	19	22	21	22	21	25
	18	16	19	13	19	16	17	19
	18	13	17	10	17	13	16	11
LOW	20	09	20	10	19	11	19	07

GRADE 6

	READ		MATH		SCI		SOC. ST	
	87	88	87	88	87	88	87	88
HIGH	21	32	22	58	24	64	23	50
	20	20	19	20	20	16	20	20
	19	18	19	12	18	09	18	13
	19	17	18	07	18	06	20	10
LOW	21	13	22	03	20	05	19	07

GRADE 8

	READ		MATH		SCI		SOC.ST	
	87	88	87	88	87	88	87	88
HIGH	25	45	28	60	27	59	28	50
	20	23	22	17	19	17	20	20
	20	15	19	11	22	11	19	13
	17	11	16	08	16	06	15	09
LOW	18	06	15	04	16	07	18	08

GRADE 10

	READ		MATH		SCI		SOC.ST	
	87	88	87	88	87	88	87	88
HIGH	24	34	25	33	25	50	24	32
	21	20	21	22	21	18	21	22
	18	18	19	18	19	12	19	19
	19	16	17	15	17	09	18	15
LOW	18	12	18	12	18	11	18	12

Table 6 -- Average percent of students in each fifth of the distributions of the 109 "academic achievement" districts.